REMARKS

Claims 1-10 stand in the present application. Reconsideration and favorable action is respectfully requested in view of the following remarks.

In the Office Action, the Examiner has rejected claims 1-10 under 35 U.S.C. § 102(b) as being anticipated by Choi. Applicants respectfully traverse the Examiner's § 102 rejection of the claims.

In Applicants' invention, as clearly required by independent claims 1, 3, 6 , 8 and 10, the semantic similarity of a sequence of words from the document set is determined and acted upon. For example, claim 1 requires:

> 1. A method for determining the semantic similarity of words in a plurality of words selected from a set of one or more documents, for use in the retrieval of information in an information system, comprising the steps of:
>
> (i) for each word of said plurality of words:
>
> (a) identifying, in documents of said set of one or more documents, word sequences comprising the word and a predetermined number of other words;
>
> (b) calculating a relative frequency of occurrence for each distinct word sequence among word sequences containing the word; and
>
> (c) generating a fuzzy set comprising, for word sequences containing the word, corresponding fuzzy membership values calculated from the relative frequencies determined at step (b); and
>
> (ii) calculating and storing, for each pair of words of said plurality of words, using respective fuzzy sets generated at step (i), a probability that the first word of the pair is semantically suitable as a replacement for the second word of the pair.

See claim 1 (emphasis supplied). Support for claim 1 can be found in the present application at paragraph [0050] and Figure 1 (reference numeral 105). Moreover, this

process is described in more detail in the present application with reference to Figure 2 (steps 200-225) and in paragraphs [0052] to [0082].

In summary, the text of the document set is stemmed, optionally after the exclusion of the most and least common words from the document set (see http://en.wikipedia.org/wiki/Word_stemming for an overview of stemming) and the resulting word output is analyzed to determine a number of n-grams. Paragraphs [0055] - [0063] provide an example of how four sentences may be analyzed to generate a number of 3-grams (that is an n-gram for the case where n=3).

Choi simply does not teach or suggest the requirements of the present claims involving integers (i) (a)-(c) of independent claims 1, 3, 6 , 8 & 10. Instead, Choi discloses an entirely different approach to the problem of ranking documents and document contents. Rather than extracting key words from the document sets, as is disclosed and claimed in Applicants' invention, Choi discloses that the user provides a set of query words. See Choi at S10 of paragraph [0051] and Figure 2, and the text of Figure 4 ("QUERY WORD GIVEN BY USER") and associated paragraphs [0086] to [0087]. Thus the method disclosed by Choi is limited by the query word(s) provided by the user; a user who has less experience of formulating a query word set, or who has less understanding of the content of document sets, will not obtain the same results as a more skilful, experienced or knowledgeable user will be able to obtain.

The Examiner alleges that integers (i) (a)-(c) of independent claims 1, 3, 6 , 8 and 10 are taught by Choi at paragraphs [0002], [0027] and/or [0143]. See Office Action at pages 2-3. However, none of the cited paragraphs of Choi discloses or even suggests: "identifying, . . . , word sequences comprising the word and a predetermined

number of other words; calculating a relative frequency of occurrence for each distinct word sequence . . . ; and generating a fuzzy set comprising, for word sequences containing the word, corresponding fuzzy membership values calculated from the relative frequencies . . . " as required by the present claims. Indeed, paragraph [0002] states that Choi relates "to a method of order ranking document clusters using entrophy data and Bayesian self-organizing feature maps." There is simply no mention anywhere in this cited paragraph of "identifying, . . . , word sequences comprising the word and a predetermined number of other words" as required by the present claims.

Cited paragraph [0027] also states that Choi provides "a method of order-ranking document clusters using entrophy data and Bayesian SOM [self-organizing feature maps]." There is simply no mention anywhere in this cited paragraph of "calculating a relative frequency of occurrence for each distinct word sequence . . . " as required by the present claims.

Finally, cited paragraph [0143] states that Choi "adopts a clustering method where documents are clustered by statistical similarity, i.e., standardized distance between the two documents." There is simply no mention anywhere in this cited paragraph of "generating a fuzzy set comprising, for word sequences containing the word, corresponding fuzzy membership values calculated from the relative frequencies . . . " as required by the present claims.

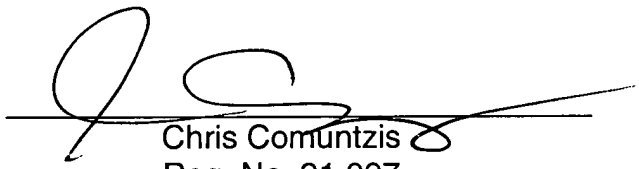
It is therefore respectfully submitted that the present claims patentably define over the cited reference, as Choi does not disclose the identification of word sequences as set out in the above noted claim integers (i)(a)-(c). A key advantage of Applicants' invention is that it obviates the need for a human user to provide a set of query words,

as required by Choi. Applicants' invention thus enables an automatic process which is independent of operator skill, experience or knowledge. Choi simply does not teach or even suggest such a solution and, thus, it is believed that the present claims patentably over Choi.

Therefore, in view of the above remarks, it is respectfully requested that the application be reconsidered and that all of claims 1-10, standing in the application, be allowed and that the case be passed to issue. If there are any other issues remaining which the Examiner believes could be resolved through either a supplemental response or an Examiner's amendment, the Examiner is respectfully requested to contact the undersigned at the local telephone exchange indicated below.

Respectfully submitted,

NIXON & VANDERHYE P.C.

By: 

Chris Comuntzis
Reg. No. 31,097

CC:lmr
901 North Glebe Road, 11th Floor
Arlington, VA 22203-1808
Telephone: (703) 816-4000
Facsimile: (703) 816-4100